Parametric continuity of stationary distributions

Cuong Le Van, John Stachurski

Parametric Continuity of Stationary Distributions ⋆

Cuong Le Van

CERMSEM, Université Paris 1 Panthéon-Sorbonne, 106-112 Boulevard de l'Hopital, France

John Stachurski

Department of Economics, The University of Melbourne, VIC 3010, Australia

Abstract

For Markovian economic models, long-run equilibria are typically identified with the stationary (invariant) distributions generated by the model. In this paper we provide new sufficient conditions for continuity in the map from parameters to these equilibria. Several existing results are shown to be special cases of our theorem.

Journal of Economic Literature Classifications C61, C62

Key words: Markov processes, stochastic dynamics, parametric continuity

1 Introduction

In economic dynamics, one frequently considers economies where the sequence of state variables $(X_t)_{t=0}^{\infty}$ is stationary. Here X_t is a vector of endogenous and exogenous variables, jointly following a Markov process generated by some underlying model. In the Markov case, stationarity reduces to the existence of a "stationary distribution" μ, such that if X_t has law μ, then so does X_{t+j} for all $j \in \mathbb{N}$. If such a μ exists then it naturally becomes the focus of equilibrium

⋆ We are indebted to an anonymous referee and the associate editor for very helpful comments. The second author is grateful for financial support from Australian Research Council Grant DP0557625.

Email addresses: `Cuong.Le-Van@univ-paris1.fr` (Cuong Le Van), `jstac@unimelb.edu.au` (John Stachurski).

Preprint submitted to Elsevier Science *1 June 2005*

analysis. For example, if μ is also unique and has some stability properties, then a law of large numbers result often holds, in which case sample moments from the series $(X_t)_{t=0}^{\infty}$ can be identified with integrals of the relevant functions with respect to the stationary distribution μ.

Typically, the underlying laws which drive the process $(X_t)_{t=0}^{\infty}$ depend on a vector of parameters, which may for example be policy instruments, or regression coefficients to be estimated from the data. In this case the parameters themselves determine the stationary distribution. The study of how this distribution varies with the parameters is a stochastic analogue of standard comparative dynamics. Our paper investigates conditions under which the functional relationship between parameters and stationary distributions is continuous.

Parametric continuity of stationary distributions is a component of various problems in estimation, simulation, numerical dynamic programming and economic theory. A well-known example is the Simulated Moments Estimator of Duffie and Singleton (1993), who require parametric continuity in order to establish consistency and other asymptotic properties of their estimators. More recently, Fernández-Villaverde, Rubio-Ramírez and Santos (2004) give conditions for convergence of the likelihood function for many numerical approximations of dynamic macroeconomic models. Again, parametric continuity is central to their study. Other important papers related to the accuracy of numerical approximation include Santos and Vigo-Aguiar (1998) and Santos and Peralta-Alva (2003).

In this paper, we use Berge's Theorem of the Maximum to provide a new parametric continuity result. The basic idea is as follows. Suppose that stationary distributions can be identified as the fixed points of a certain operator P_θ mapping distributions into distributions, where $\theta \in \Theta$ is a parameter. If we can furnish a metric ϱ on the space of distributions, then the function $F(\theta, \mu) := -\varrho(\mu, P_\theta(\mu))$ is zero if and only if μ is stationary given θ. In fact, providing that at least one stationary distribution exists for each θ, it is clear that the set of stationary distributions and the set of maximizers of $\mu \mapsto F(\theta, \mu)$ coincide. When Berge's conditions are satisfied, his Theorem of the Maximum tells us precisely when the dependence of these maximizers on the parameters will be continuous.

The main theorem includes some well-known results as special cases. One is a result in Stokey, Lucas and Prescott (1989, Theorem 12.13) for Markov models on a compact state space. Another is due to Stenflo (2001), who proves parametric continuity for noncompact state spaces when the transition rule is contracting on average. His assumptions are shown to imply the conditions of our theorem whenever the closed and bounded subsets of the state space are compact (as is the case, for example, with $(\mathbb{R}^n, \|\cdot\|)$). We also provide a new result which is another special case of the main theorem, and should prove useful in applications. This claim is illustrated using a simple growth model.

2 Set Up

Let $\mathscr{P}(S)$ be the collection of probabilities on $(S, \mathscr{B}(S))$, where S is any separable, completely metrizable topological space, and $\mathscr{B}(S)$ is its Borel sets. Let $\mathscr{M}(S)$ be the linear space of finite signed measures on $(S, \mathscr{B}(S))$, and let $bC(S)$ be the bounded continuous real valued functions on S. For $\mu \in \mathscr{M}(S)$ and $h \in bC(S)$ we use the symmetric notation $\langle \mu, h \rangle = \langle h, \mu \rangle$ to denote $\int_S h d\mu$. Let $w(\mathscr{M}(S), bC(S))$ be the weak topology on $\mathscr{M}(S)$ generated by the set of linear functionals $\mu \mapsto \langle \mu, h \rangle$, $h \in bC(S)$, in the usual way (see, e.g., Stokey, Lucas and Prescott, Chapter 12), and let $w(\mathscr{P}(S), bC(S))$ be the relative topology on $\mathscr{P}(S)$.

In the proofs we use the Fortet-Mourier metrization of $w(\mathscr{P}(S), bC(S))$: Let d be any distance function which metrizes the topology on S. Let $BL(S, d)$ be the collection of bounded Lipschitz functions on (S, d). This space is given the norm

$$\|h\|_{BL} := \sup_{x \in S} |h(x)| + \sup_{x \neq y} \frac{|h(x) - h(y)|}{d(x, y)}. \tag{1}$$

Now set $\varrho_{FM}(\mu, \nu) := \sup |\langle \mu - \nu, h \rangle|$, where the supremum is over all $h \in BL(S, d)$ with $\|h\|_{BL} \leq 1$. Given that S is separable, the function ϱ_{FM} so defined is known to metrize $w(\mathscr{P}(S), C_b(S))$ (cf., e.g., Dudley 2002, Theorem 11.3.3).

A stochastic kernel (or transition probability function) on S is a map $P \colon S \times \mathscr{B}(S) \to [0, 1]$ with the property that $x \mapsto P(x, B)$ is Borel measurable for each $B \in \mathscr{B}(S)$, and $B \mapsto P(x, B)$ is an element of $\mathscr{P}(S)$ for each $x \in S$. We set $Ph(x) := \int_S h(y) P(x, dy)$ for real valued h on S where this integral is defined. In addition, for $\mu \in \mathscr{M}(S)$, we write μP for the element of $\mathscr{M}(S)$ defined by $(\mu P)(B) := \int P(x, B) \mu(dx)$. Thus, P is an operator which acts on functions to the right and measures to the left. [1]

It can easily be shown that $h \mapsto Ph$ is a positive (i.e., increasing) linear operator on $bC(S)$, as is $\mu \mapsto \mu P$ on $\mathscr{M}(S)$. Clearly $P \mathbb{1}_S = \mathbb{1}_S$. Also, we have $\langle \mu P, h \rangle = \langle Ph, \mu \rangle$ for all $h \in bC(S)$ and all $\mu \in \mathscr{P}(S)$. [2] For $x \in S$ we use δ_x

[1] This notation is quite standard. See, for example, the classic monograph of Meyn and Tweedie (1993).

[2] In other words, the two operators are adjoint. See Stokey, Lucas and Prescott (1989, Theorem 8.3).

to denote the probability with unit mass on x. We let P^t denote t compositions of P with itself.[3]

Given P, a stationary or invariant distribution is a $\mu \in \mathscr{P}(S)$ such that $\mu P = \mu$. A function $V\colon S \to [0, \infty)$ is called a Lyapunov function (or simply Lyapunov) if it is continuous and all sublevel sets $\{x \in S : V(x) \leq a\}$ are compact.[4] Let $\mathscr{L}(S)$ be the set of Lyapunov functions on S. Finally, a subset Q of $\mathscr{P}(S)$ is called tight if, for all $\varepsilon > 0$, there is a compact $K \subset S$ such that $\sup_{\mu \in Q} \mu(S \setminus K) \leq \varepsilon$.

3 Results

Our starting point is a parameter space Θ and a family of stochastic kernels $\{P_\theta : \theta \in \Theta\}$. Here Θ is an arbitrary topological space. Let N denote any subset of Θ. Define $\Lambda(\theta) := \{\mu \in \mathscr{P}(S) : \mu = \mu P_\theta\}$.

Assumption 3.1 *$N \times \mathscr{P}(S) \ni (\theta, \mu) \mapsto \mu P_\theta \in \mathscr{P}(S)$ is continuous.*

Assumption 3.2 *For each $\theta \in N$, there is a $V \in \mathscr{L}(S)$ and $x \in S$ such that $\liminf_{t \to \infty} P_\theta^t V(x) < \infty$.*

The following existence result is immediate from Meyn and Tweedie (1993, Proposition 12.1.3).

Lemma 3.1 *If Assumptions 3.1 and 3.2 hold, then $\Lambda(\theta)$ is nonempty for all $\theta \in N$.*

Parametric continuity is a classic problem of interchanging orders of limits. In such situations a degree of uniformity is usually necessary. The next assumption is a uniform compactness requirement. To state it we use the following notation: For $W \in \mathscr{L}(S)$ and $M \in \mathbb{N}$ define $\Gamma(W, M) := \{\mu \in \mathscr{P}(S) : \int W d\mu \leq M\}$.

Assumption 3.3 *There exists a $W \in \mathscr{L}(S)$ and an $M \in \mathbb{N}$ such that $\Lambda(\theta) \subset \Gamma(W, M)$ for all $\theta \in N$.*[5]

[3] It is well-known that $\delta_x P^t$ is the marginal distribution of X_t given that $X_0 \equiv x \in S$, and $(X_t)_{t=0}^\infty$ follows the Markov process defined by P; while $P^t h(x)$ is the expectation of $h(X_t)$ conditional on $X_0 \equiv x$. See, for example, Stokey, Lucas and Prescott (1989, p. 213).

[4] For example, if S is compact then every continuous nonnegative real function is Lyapunov. Alternatively, if d metrizes the topology on S and the closed bounded subsets of (S, d) are compact, then $V(x) = d(x, x_0)$ is Lyapunov for each $x_0 \in S$.

[5] In applying Assumptions 3.2 and 3.3 we make use of the following result: If

Using it we can present our main result:

Theorem 3.1 *If Assumptions 3.1–3.3 hold for some $N \subset \Theta$, then the correspondence $\theta \mapsto \Lambda(\theta)$ is nonempty, compact valued, and upper hemicontinuous on N.*

Proof. Define $F(\theta, \mu) := -\varrho_{FM}(\mu, \mu P_\theta)$.[6] Taking W and M as given in Assumption 3.3, set $H(\theta) := \operatorname{argmax}_{\mu \in \Gamma(W,M)} F(\theta, \mu)$. Note that $\mu \in \Lambda(\theta)$ iff $F(\theta, \mu) = 0$. Also, by Assumption 3.1, the function F is continuous on $N \times \mathscr{P}(S)$. Furthermore, $\Gamma(W, M)$ is compact (see the comments in footnote 5) and nonempty (by Lemma 3.1 and Assumption 3.3). Berge's Theorem of the Maximum (Aliprantis and Border, 1999, p. 539) then implies that $\theta \mapsto H(\theta)$ is upper hemicontinous on N. Finally, observe that $H(\theta) = \Lambda(\theta)$ for all $\theta \in N$, because $\Lambda(\theta) \subset \Gamma(W, M)$ by Assumption 3.3, and $\Lambda(\theta)$ is nonempty (recall Lemma 3.1).

Remark 3.1 *For example, if there is a unique fixed point μ_θ for each $\theta \in N$, then $\theta \mapsto \mu_\theta$ is continuous on N.*

4 Existing Applications

In this section we show how some seemingly unrelated existing results can be derived from Theorem 3.1.

4.1 Compact State

First, consider the compact state space result of Stokey, Lucas and Prescott (1989, Theorem 12.13), which is apparently due to R.E. Manuelli:

Theorem 4.1 *Let S be compact. If Assumption 3.1 holds for some $N \subset \Theta$ and $\Lambda(\theta)$ is single valued, then $\theta \mapsto \Lambda(\theta)$ is continuous on N.*

This result is immediate from Theorem 3.1: Set $V = W = 0$ everywhere on S and let $M = 0$ in Assumptions 3.2 and 3.3.

$V \in \mathscr{L}(S)$, $M \in \mathbb{N}$, and $Q \subset \mathscr{P}(S)$ with $\sup_{\mu \in Q} \int V d\mu \leq M$ then Q is tight. (The proof is not difficult. See Meyn and Tweedie, 1993, Lemma D.5.3.) The closure of Q is then $w(\mathscr{P}(S), bC(S))$-compact by Prohorov's theorem.

[6] The metric ϱ_{FM} was defined above. In fact any distance function which metrizes the topology on S will do.

Even though this theorem is quite straightforward, it is not always easy to check Assumption 3.1 in applications. For example, the joint continuity of $(\theta, \mu) \mapsto \mu P_\theta$ is more difficult to check that the requirement that $\mu \mapsto \mu P_\theta$ and $\theta \mapsto \mu P_\theta$ are continuous for each θ and μ respectively. Moreover, the immediate object of interest in economic studies is usually a stochastic difference equation, rather than a stochastic kernel. Finally, in much of applied macroeconomics the state space is not compact. Below we discuss results which address some of these concerns.

4.2 Average Contractions

In this section we review the results of Stenflo (2001, Theorem 2). Suppose that $S = (S, d)$ is boundedly compact.[7] In this case it turns out that his parametric continuity theorem is also a special case of Theorem 3.1.[8] To state his theorem, let $(Z, \mathscr{Z})$ be an arbitrary measurable space, and let $\mathscr{P}(Z)$ be the probabilities on $(Z, \mathscr{Z})$. Stenflo considers the stochastic recursive model

$$X_{t+1} = T_\theta(X_t, \xi_{t+1}), \quad \text{where} \quad \xi_t \sim \psi_\theta \in \mathscr{P}(Z), \quad \forall t \in \mathbb{N}. \tag{2}$$

Here T_θ is a measurable function sending $S \times Z \to S$ for each $\theta \in \Theta$, and $(\xi_t)_{t=1}^\infty$ is an independent sequence, all with distribution ψ_θ. For $x \in S$ and $B \in \mathscr{B}(S)$ we set $P_\theta(x, B) := \psi_\theta\{z \in Z : T_\theta(x, z) \in B\}$. Stenflo restricts attention to the case where $\Theta = (\Theta, e)$ is a metric space (e is the metric on Θ). He makes the following assumptions, where, as before, N is an arbitrary subset of Θ:

Assumption 4.1 *There exists a $\lambda \in (0, 1)$ such that, $\forall\, \theta \in N$,*

$$\int d(T_\theta(x, z), T_\theta(x', z))\psi_\theta(dz) \leq \lambda d(x, x'), \quad \forall\, x, x' \in S.$$

Assumption 4.2 *There exists an $x_0 \in S$ such that*

$$L := \sup_{\theta \in N} \int d(T_\theta(x_0, z), x_0)\psi_\theta(dz) < \infty.$$

[7] A metric space is called boundedly compact if all the closed balls are compact. The finite dimensional vector spaces are typical examples. We need bounded compactness of S to ensure that $x \mapsto d(x, x_0)$ is Lyapunov on S for all $x_0 \in S$.

[8] It should be noted, however, that Stenflo obtains rates of convergence. Rates are useful for deriving error bounds in computational problems. See also Santos and Peralta-Alva (2004, Theorem 4.2). In contrast, Theorem 3.1 cannot be used to derive rates.

It is known (see, e.g., Stenflo, 2001, Theorem 1) that

Lemma 4.1 *If Assumptions 4.1 and 4.2 hold, then P_θ has a unique stationary distribution $\mu_\theta \in \mathscr{P}(S)$ for each $\theta \in N$. Moreover, for each $x \in S$ and $\theta \in N$ we have $\delta_x P_\theta^t \to \mu_\theta$ as $t \to \infty$.*

To derive parametric continuity he requires in addition:

Assumption 4.3 *There exists a function δ mapping $[0, \infty)$ to itself such that $\delta(x) \to 0$ when $x \to 0$, and*

$$\sup_{z \in Z} \sup_{x \in S} d(T_\theta(x,z), T_{\theta'}(x,z)) \leq \delta(e(\theta, \theta')), \quad \forall\, \theta, \theta' \in N.$$

Assumption 4.4 *The map $N \ni \theta \mapsto \psi_\theta \in \mathscr{P}(Z)$ is continuous with respect to the total variation norm topology on $\mathscr{P}(Z)$.*

Theorem 4.2 (Stenflo) *Let μ_θ be as in Lemma 4.1. If Assumptions 4.1–4.4 all hold, then $\theta \to \mu_\theta$ is continuous on N.*

When S is boundedly compact this turns out to be a special case of Theorem 3.1:

Proposition 4.1 *If S is boundedly compact, then Assumptions 4.1—4.4 imply Assumptions 3.1–3.3, with $V(x) = W(x) = d(x, x_0)$ and $M = L/(1-\lambda)$.*

Proof. First we verify Assumption 3.1. To do so, pick any (θ, μ) in $N \times \mathscr{P}(S)$, and any sequence $(\theta_n, \mu_n)_{n=1}^\infty \subset N \times \mathscr{P}(S)$ converging to (θ, μ). Let $h \in BL(S,d)$, $\|h\|_{BL} \leq 1$, and consider

$$|\langle \mu_n P_{\theta_n} - \mu P_\theta, h \rangle| = |\langle P_{\theta_n} h, \mu_n \rangle - \langle P_\theta h, \mu \rangle|, \tag{3}$$

which is dominated by

$$|\langle P_{\theta_n} h, \mu_n \rangle - \langle P_{\theta_n} h, \mu \rangle| + |\langle P_{\theta_n} h, \mu \rangle - \langle P_\theta h, \mu \rangle|. \tag{4}$$

To bound the first term in (4), we make use of the following elementary observations. First, if $g \in BL(S,d)$ and $\|g\|_{BL} \leq r$, then $\|(2r)^{-1} g\|_{BL} \leq 1$; from which we can see that if μ and $\mu' \in \mathscr{P}(S)$, and $g \in BL(S,d)$ with $\|g\|_{BL} \leq r$, then $|\langle \mu - \mu', g \rangle| \leq 2r \varrho_{FM}(\mu, \mu')$. Finally, taking h as given, suppose we define $g_n(x) := P_{\theta_n} h(x)$. Evidently $|g_n| \leq |h|$, and

$$\begin{aligned} |g_n(x) - g_n(x')| &= \left| \int h(T_{\theta_n}(x,z)) \psi_{\theta_n}(dz) - \int h(T_{\theta_n}(x',z)) \psi_{\theta_n}(dz) \right| \\ &\leq \int |h(T_{\theta_n}(x,z)) - h(T_{\theta_n}(x',z))| \psi_{\theta_n}(dz) \\ &\leq \int d(T_{\theta_n}(x,z), T_{\theta_n}(x',z)) \psi_{\theta_n}(dz). \end{aligned}$$

Assumption 4.1 now gives

$$|g_n(x) - g_n(x')| \leq \lambda d(x, x'), \quad \forall x, x' \in S, \ \forall n \in \mathbb{N}. \tag{5}$$

It follows that $g_n \in BL(S, d)$ and $\|g_n\|_{BL} \leq 2$ for all n.

From these observations bounding the first term in (4) is now easy. We have

$$|\langle P_{\theta_n} h, \mu_n \rangle - \langle P_{\theta_n} h, \mu \rangle| = |\langle g_{\theta_n}, \mu_n \rangle - \langle g_{\theta_n}, \mu \rangle| \leq 4\varrho_{FM}(\mu_n, \mu). \tag{6}$$

Next, we consider the second term in (4). Clearly

$$\begin{aligned} &|\langle P_{\theta_n} h, \mu \rangle - \langle P_\theta h, \mu \rangle| \\ &\qquad \leq \int \left| \int h(T_{\theta_n}(x, z)) \psi_{\theta_n}(dz) - \int h(T_\theta(x, z)) \psi_\theta(dz) \right| \mu(dx). \end{aligned}$$

Consider the term inside the absolute value symbols. It is dominated by

$$\begin{aligned} &\left| \int h(T_{\theta_n}(x, z)) \psi_{\theta_n}(dz) - \int h(T_\theta(x, z)) \psi_{\theta_n}(dz) \right| \\ &\qquad + \left| \int h(T_\theta(x, z)) \psi_{\theta_n}(dz) - \int h(T_\theta(x, z)) \psi_\theta(dz) \right|. \end{aligned} \tag{7}$$

From Assumption 4.3, the first term in this sum is bounded above by

$$\begin{aligned} &\int |h(T_{\theta_n}(x, z)) - h(T_\theta(x, z))| \psi_{\theta_n}(dz) \\ &\qquad \leq \int d(T_{\theta_n}(x, z), T_\theta(x, z)) \psi_{\theta_n}(dz) \leq \delta(e(\theta_n, \theta)). \end{aligned} \tag{8}$$

Since $|h| \leq 1$, the second term in the sum (7) is bounded above by $\|\psi_{\theta_n} - \psi_\theta\|$, where $\|\cdot\|$ is the total variation norm on $\mathscr{P}(Z)$.

$$\therefore \quad |\langle P_{\theta_n} h, \mu \rangle - \langle P_\theta h, \mu \rangle| \leq \delta(e(\theta_n, \theta)) + \|\psi_{\theta_n} - \psi_\theta\|. \tag{9}$$

Combining (3), (4), (6) and (9) gives

$$|\langle \mu_n P_{\theta_n} - \mu P_\theta, h \rangle| \leq 4\varrho_{FM}(\mu_n, \mu) + \delta(e(\theta_n, \theta)) + \|\psi_{\theta_n} - \psi_\theta\|.$$

Since h was an arbitrary element of the unit ball of $BL(S, d)$, we have

$$\varrho_{FM}(\mu_n P_{\theta_n}, \mu P_\theta) \leq 4\varrho_{FM}(\mu_n, \mu) + \delta(e(\theta_n, \theta)) + \|\psi_{\theta_n} - \psi_\theta\|.$$

The required continuity of $(\theta, \mu) \mapsto \mu P_\theta$ is now verified by Assumptions 4.3 and 4.4.

Next we prove Assumptions 3.2 and 3.3 with $V(x) = W(x) = d(x, x_0)$ and $M = L/(1-\lambda)$. Bounded compactness of S implies that $V \in \mathscr{L}(S)$. We have

$$\begin{aligned}
P_\theta V(x) &= \int V(T_\theta(x,z))\psi_\theta(dz) \\
&= \int d(T_\theta(x,z), x_0)\psi_\theta(dz) \\
&\leq \int d(T_\theta(x,z), T_\theta(x_0,z))\psi_\theta(dz) + \int d(T_\theta(x_0,z), x_0)\psi_\theta(dz) \\
&\leq \lambda V(x) + L.
\end{aligned}$$

Since P_θ is positive, linear, and $P_\theta \mathbb{1}_S = \mathbb{1}_S$, iterating gives

$$P_\theta^t V(x) \leq \lambda^t V(x) + \lambda^{t-1} L + \lambda^{t-2} L + \cdots + L.$$

This and the fact that λ and L are independent of θ provides the uniform bound

$$\sup_{\theta \in N} \sup_{t \geq 1} P_\theta^t V(x) \leq V(x) + \frac{L}{1-\lambda}.$$

In particular, for $x = x_0$ we get $\sup_{\theta \in N} \sup_{t \geq 1} P_\theta^t V(x_0) \leq L/(1-\lambda)$, which verifies Assumption 3.2.

Now let $V_n := V \wedge n$ be the n-th truncation of V, and let μ_θ be the stationary distribution corresponding to θ. Since $V_n \in bC(S)$, $\forall n \in \mathbb{N}$, Lemma 4.1 and the definition of convergence in $w(\mathscr{P}(S), bC(S))$ imply that

$$\lim_t P_\theta^t V_n(x_0) = \int V_n d\mu_\theta. \tag{10}$$

Also, since P_θ and hence P_θ^t are positive operators, we have $P_\theta^t V_n(x_0) \leq P_\theta^t V(x_0)$, which in turn is bounded by $L/(1-\lambda)$. The Monotone Convergence Theorem now gives

$$\int V d\mu_\theta = \lim_n \int V_n d\mu_\theta = \lim_n \lim_t P_\theta^t V_n(x_0) \leq \frac{L}{1-\lambda}, \quad \forall \theta \in N.$$

Assumption 3.3 is therefore satisfied with $W(x) = V(x) = d(x, x_0)$ and $M = L/(1-\lambda)$.

5 Further Applications

Next we develop a new application of Theorem 3.1, which extends Stenflo's results in Section 4.2. So let S and Z be as in that section (although S need not be boundedly compact), and consider the model

$$X_{t+1} = T_\theta(X_t, \xi_{t+1}), \text{ where } \xi_t \sim \psi \in \mathscr{P}(Z), \ \forall t \in \mathbb{N}. \tag{11}$$

As before, $T_\theta \colon S \times Z \to S$ is measurable, $(\xi_t)_{t=1}^\infty$ is IID, and N is an arbitrary subset of (Θ, e). Set $P_\theta(x, B) := \psi\{z \in Z : T_\theta(x, z) \in B\}$.

First, we wish to weaken Stenflo's Assumption 4.3, which is too restrictive in some applications (see the growth model example below). The following assumption is clearly weaker.

Assumption 5.1 *$N \ni \theta \mapsto T_\theta(x, z) \in S$ is continuous for each pair $(x, z) \in S \times Z$.*

We wish also to relax Stenflo's Assumption 4.1, which requires that the law of motion is contracting on average. This may or may not be satisfied in applications. For example, if we take $S = Z = \mathbb{R}$, $d(x, y) = |x - y|$, and law of motion $X_{t+1} = g_\theta(X_t) + \xi_{t+1}$, then Assumption 4.1 requires that g_θ has slope with absolute value less than one everywhere on $\mathbb{R}$. We wish to assume only that g_θ be locally Lipschitz. This will be the case if, for example, g_θ is either Lipschitz or continuously differentiable.

Assumption 5.2 *For each compact $C \subset S$, there is a $K < \infty$ s.t.*

$$\int d(T_\theta(x, z), T_\theta(x', z))\psi(dz) \le K d(x, x'), \quad \forall x, x' \in C, \ \ \forall\, \theta \in N.$$

Finally, we require a drift condition with respect to a Lyapunov function, which has the effect of shifting probability mass towards areas of the state space where the Lyapunov function is small:

Assumption 5.3 *There exists a $V \in \mathscr{L}(S)$, $\lambda \in (0, 1)$ and $L \in [0, \infty)$ such that, $\forall \theta \in N$,*

$$P_\theta V(x) = \int V(T_\theta(x, z))\psi(dz) \le \lambda V(x) + L, \quad \forall x \in S.$$

Under these assumptions we have the following result:

Proposition 5.1 *Let $\theta \in N$. If Assumptions 5.1–5.3 hold, then $\Lambda(\theta)$ is nonempty. If $\Lambda(\theta) = \{\mu_\theta\}$, then $\theta \mapsto \mu_\theta$ is continuous on N.*

Proof. First we verify Assumption 3.1. As in the proof of Proposition 4.1, let $(\theta, \mu) \in N \times \mathscr{P}(S)$, and let $(\theta_n, \mu_n)_{n=1}^\infty \subset N \times \mathscr{P}(S)$ be a sequence converging to (θ, μ). Fix $h \in BL(S, d)$, $\|h\|_{BL} \le 1$. It is sufficient to show that (3) converges to zero as $n \to \infty$ (Dudley, 2002, Theorem 11.3.3). In fact it is sufficient to show that any subsequence has a subsubsequence converging to zero. To simplify notation we let (θ_n, μ_n) be the arbitrary subsequence.

Now fix $\varepsilon > 0$, and consider again the first term in (4). Let $g_n(x) := P_{\theta_n} h(x)$, and $g(x) := P_\theta h(x)$. By Assumption 5.1, and the Dominated Convergence Theorem g_n converges pointwise to g. Evidently $|g_n| \leq |h|$, and

$$\begin{aligned} |g_n(x) - g_n(x')| &= \left| \int h(T_{\theta_n}(x, z)) \psi(dz) - \int h(T_{\theta_n}(x', z)) \psi(dz) \right| \\ &\leq \int |h(T_{\theta_n}(x, z)) - h(T_{\theta_n}(x', z))| \psi(dz) \\ &\leq \int d(T_{\theta_n}(x, z), T_{\theta_n}(x', z)) \psi(dz). \end{aligned}$$

Since for a separable and completely metrizable space S any convergent sequence in $\mathscr{P}(S)$ is tight (Dudley, 2002, Theorem 11.5.3), we can take a compact set $C \subset S$ such that $\sup_n \mu_n(S \setminus C) \leq \varepsilon$. Assumption 5.2 gives

$$|g_n(x) - g_n(x')| \leq K d(x, x'), \quad \forall x, x' \in C, \ \forall n. \tag{12}$$

Thus, restricted to C, $\{g_n\}$ is a uniformly bounded and equicontinuous sequence of functions. By the Arzelà-Ascoli Theorem, $\{g_n\}$ is precompact in the sup norm topology, and therefore has a *uniformly* convergent subsequence $\{g_{n(j)}\}$. Obviously the limit of this subsequence is g, so that, for some $J \in \mathbb{N}$, $|g_{n(j)}(x) - g(x)| \leq \varepsilon$ for all $x \in C$ and all $j \geq J$. For all such j, $\sup_n \mu_n(S \setminus C) \leq \varepsilon$ implies

$$\begin{aligned} |\langle P_{\theta_{n(j)}} h, \mu_{n(j)} \rangle - \langle P_{\theta_{n(j)}} h, \mu \rangle| &= \left| \int g_{n(j)} d\mu_{n(j)} - \int g_{n(j)} d\mu \right| \\ &\leq \left| \int_C g_{n(j)} d\mu_{n(j)} - \int_C g_{n(j)} d\mu \right| + 2\varepsilon. \end{aligned}$$

Replacing $g_{n(j)}$ with g we get

$$|\langle P_{\theta_{n(j)}} h, \mu_{n(j)} \rangle - \langle P_{\theta_{n(j)}} h, \mu \rangle| \leq \left| \int_C g d\mu_{n(j)} - \int_C g d\mu \right| + 4\varepsilon.$$

Since g is continuous and bounded on C, and since the restriction of μ_n to C converges in $w(\mathscr{P}(S), bC(S))$ to the restriction of μ to C, the term on the right goes to zero in j.

Regarding the second term in (4), clearly it is dominated by

$$\int \int |h(T_{\theta_n}(x, z)) - h(T_\theta(x, z))| \psi(dz) \mu(dx).$$

By Assumption 5.1 and the Dominated Convergence Theorem this goes to zero in n. Assumption 3.1 is verified.

Now we argue that $\Lambda(\theta)$ is nonempty for each $\theta \in N$. An identical argument to the iterative procedure used in the proof of Proposition 4.1 yields

$$\sup_{\theta \in N} \sup_{t \geq 1} P_\theta^t V(x) \leq V(x) + \frac{L}{1-\lambda}. \tag{13}$$

Moreover, it is easy to see that Assumption 5.2 implies P_θ is Feller for each $\theta \in N$ (see Stokey, Lucas and Prescott, 1989, p. 220 for a definition). Existence of a stationary distribution μ_θ now follows from Meyn and Tweedie (1993, Proposition 12.1.3). Clearly Assumption 3.2 is also verified by (13).

It only remains to check Assumption 3.3 under the hypothesis that $\Lambda(\theta) = \{\mu_\theta\}$ is single-valued. Define from P_θ the new operator $\bar{P}_\theta$ by $\bar{P}_\theta := t^{-1} \sum_{j=1}^t P_\theta^j$. By Meyn and Tweedie (1993, Proposition 12.1.4), $\delta_x \bar{P}^t \to \mu_\theta$ as $t \to \infty$ for all $x \in S$. Repeating exactly the verification of Assumption 3.3 in Proposition 4.1, but replacing P_θ by $\bar{P}_\theta$, we can see that Assumption 3.3 also holds under the hypotheses of Proposition 5.1. The proof is done.

6 Example

Consider the following simple example. A representative household maximizes

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t (\eta \ln c_t + (1-\eta) \ln \ell_t),$$

subject to $c_t + k_{t+1} \leq A k_t^\alpha \ell_t^{1-\alpha} \varepsilon_{t+1}$, $\alpha \in (0,1)$. We take $(\varepsilon_t)_{t=1}^\infty$ as IID on $(0, \infty)$. It is well-known that the optimal accumulation policy for this model is given by $k_{t+1} = \alpha\beta A k_t^\alpha \ell^{1-\alpha} \varepsilon_{t+1}$, where ℓ is a constant depending on the parameters. Taking logs and setting $\kappa := \ln k$ and $\xi := \ln \varepsilon$ gives

$$\kappa_{t+1} = b + \alpha \kappa_t + \xi_{t+1}. \tag{14}$$

Let $\xi \sim \psi \in \mathscr{P}(\mathbb{R})$, with $\mathbb{E}|\xi| := \int |z| \psi(dz) < \infty$. Also, let $S = Z = \mathbb{R}$, and let $d(x,y) = |x-y|$. Finally, although b depends on several parameters it is sufficient for our purposes to regard it as a single parameter taking values in $\mathbb{R}$. With this convention we can take

$$\theta := (b, \alpha) \ni \mathbb{R} \times (0,1) =: \Theta,$$

and $T_\theta(\kappa, z) = b + \alpha\kappa + z$. For this model we cannot apply Stenflo's parametric continuity result, because Assumption 4.3 is not satisfied. To see this, take

$\theta = (b, \alpha)$ and $\theta' = (b', \alpha')$ with $\alpha \neq \alpha'$. Then

$$\sup_{\kappa \in S} d(T_\theta(\kappa, z), T_{\theta'}(\kappa, z)) = \sup_{\kappa \in S} |b + \alpha\kappa + z - b' - \alpha'\kappa - z|$$
$$\leq |b - b'| + |\alpha - \alpha'| \sup_{\kappa \in S} |\kappa| = \infty.$$

However, Proposition 5.1 is easy to apply. Let N be any open subset of Θ with compact closure $\bar{N} \subset \Theta$. By Lemma 4.1, (14) has one and only one stationary distribution μ_θ for each $\theta \in N$, so to prove that $N \ni \theta \mapsto \mu_\theta \in \mathscr{P}(S)$ is continuous we need only verify that Assumptions 5.1–5.3 hold on N.

Assumption 5.1 is trivial, as is Assumption 5.2, because for all $\theta \in N$ we have

$$d(T_\theta(\kappa, z), T_\theta(\kappa', z)) = |b + \alpha\kappa + z - b - \alpha\kappa' - z| = \alpha|\kappa - \kappa'| \leq d(\kappa, \kappa').$$

Regarding Assumption 5.3, let $V(x) := |x|$, which is clearly Lyapunov on $\mathbb{R}$. Since $\bar{N}$ is a compact subset of $\Theta = \mathbb{R} \times (0, 1)$, there is a $\lambda < 1$ and an $L_0 < \infty$ such that $\alpha \leq \lambda$ and $|b| \leq L_0$ for all $(b, \alpha) \in N$. Setting $L := L_0 + \mathbb{E}|\xi|$, we get

$$\int V(T_\theta(\kappa, z))\psi(dz) = \int |b + \alpha\kappa + z|\psi(dz)$$
$$\leq \alpha|\kappa| + |b| + \mathbb{E}|\xi| \leq \lambda V(\kappa) + L.$$

As a result, Assumptions 5.1–5.3 are all verified, Proposition 5.1 applies, and $\theta \mapsto \mu_\theta$ is continuous on N.

References

[1] Aliprantis, C. D. and K. C. Border (1999): *Infinite Dimensional Analysis,* Second Edition, Springer-Verlag, New York.

[2] Dudley, R. M. (2002): *Real Analysis and Probability,* Cambridge Studies in Advanced Mathematics No. 74, Cambridge University Press.

[3] Fernández-Villaverde, J., J.F. Rubio-Ramírez and M.S. Santos (2004): "Convergence Properties of the Likelihood of Computed Dynamic Models," PIER Working Paper 04-034.

[4] Meyn, S. P. and Tweedie, R. L. (1993): *Markov Chains and Stochastic Stability,* Springer-Verlag: London.

[5] Santos, M. S. and A. Peralta-Alva (2004): "Accuracy of Simulations for Stochastic Dynamic Models." Manuscript.

[6] Santos, M.S. and J. Vigo-Aguiar (1998): "Analysis of a Numerical Dynamic Programming Algorithm Applied to Economic Models," *Econometrica,* 66(2), 409–426.

[7] Stenflo, O. (2001): "Ergodic Theorems for Markov Chains Represented by Iterated Function Systems," *Bulletin of the Polish Academy of Sciences: Mathematics,* 49, 27–43.

[8] Stokey, N. L., R. E. Lucas and E. C. Prescott (1989): *Recursive Methods in Economic Dynamics* Harvard University Press, Massachusetts.